

Effect of Dropout on the Efficiency of Ds - Optimal Designs for Linear Mixed Models

By

Richard Kinai

Submitted to the graduate degree program in Psychology
and the Graduate Faculty of the University of Kansas in partial fulfillment of the
requirements for the degree of Master of Arts.

Chairperson: Wei Wu, PhD.

Pascal R. Deboeck, PhD.

Matthew R. Reynolds, PhD.

Date Defended: June 30, 2015

The Thesis Committee for Richard Kinai

certifies that this is the approved version of the following thesis:

Effect of Dropout on the Efficiency of Ds - Optimal Designs for Linear Mixed Models

Chairperson: Wei Wu, PhD.

Date approved: June 30, 2015

Abstract

Optimal designs are a class of experimental designs that are efficient with respect to some statistical criterion. Two types of optimal designs are considered in the study. D-optimal designs are designs that minimize the generalized variance of a model's estimated parameters. Ds-optimal designs are a class of D-optimal experimental designs that are useful when the researcher is interested in estimating a subset of parameters in a given model. For a specific parameter, Ds-optimal designs would be more efficient than D-optimal designs. Although the loss in efficiency of D-optimal designs relative to Ds-optimal designs have been examined in the past literature, past research did not consider the cases where there are missing observations.

Given that missing observations are ubiquitous in longitudinal studies due to dropout, the current study examines the loss in efficiency when D-optimal designs are used instead of Ds-optimal designs for data with missing observations. Results indicate that in general, location of Ds-optimal design points with dropout will shift closer towards the location of the D-optimal designs with complete data, compared to D-optimal design points with dropout. The D-optimal design with complete data corresponds with the smallest variance covariance matrix. For the data with dropout, the variance covariance matrix of the Ds-optimal design is closer in size to that of D-optimal design with complete data compared to that of D-optimal design with dropout. For both designs with dropout, efficiency loss is moderate.

Keywords Ds-optimal designs, missing observations, efficiency loss, linear mixed models.

Table of Contents

| | |
|--|----|
| Introduction..... | 1 |
| SECTION 1: Linear Mixed Models..... | 3 |
| SECTION 2: Missing Data | 4 |
| SECTION 3: Introduction to Optimal Designs | 5 |
| SECTION 4: Optimal Designs and response probability functions..... | 10 |
| SECTION 5: Method..... | 12 |
| SECTION 6: Results..... | 16 |
| SECTION 7: Conclusions..... | 18 |
| References..... | 21 |
| Figures..... | 23 |
| Table | 30 |

Introduction

Longitudinal studies in general tend to encounter high dropout rates besides being costly. There is evidence that missing observations reduce the efficiency of longitudinal experiments noticeably (Ortega-Azurduy et al., 2008). Designing a study to take into account missingness is key in ensuring efficiency of target statistical tests. In fact, a proper design of a scientific study is more important than the specific techniques used in the analysis. A poorly designed study or a botched experiment may often not be salvaged, even with the most sophisticated analysis (Cox, 1958). Increasing costs and limited funding necessitate the need to consider the use of optimal designs in research.

Optimal designs have not been popular in psychological research (Holling, 2013). Several reasons have been used to explain why this has been the case. Inaccessible literature (requiring a strong statistical background), common software packages lacking capabilities to generate optimal designs, and low cost experiments in psychology (hence no motivation to thoroughly plan to cut cost), are some of the reasons. Another reason is that design issues have been treated lightly in the curriculum with data analysis being the main emphasis (Wong & Berger, 2009).

However, there is a growing demand for optimal designs in psychological research in recent years. In educational testing, due to large-scale assessments, optimal designs are very valuable. An example is the study on “Trends in International Mathematics and Science” (TIMSS; Mullis, Martin, Ruddock, O’Sullivan, & Preuschoff, 2009), where the large sample sizes needed may be kept at a minimum by using optimal designs. Recent years have seen neuroimaging methods like fMRI become standard procedure in psychological research. Due to their complexity and high cost, optimal designs may reduce the time and cost involved (Maus &

van Breukelen, 2013). As a result of developments in software and more applied literature, there is a growth body of research in optimal designs in areas like survival analysis (Moerbeek, Jozwiak, 2013), item response theory (Berger, 1994), and clinical trials (Haines et. al., 2003).

From a practical standpoint, each psychological study may benefit from optimal designs because the reduction of observations without losing precision is often considerable. Besides financial constraints, there are ethical and practical reasons that justify the need to use optimal designs (McClelland, 1997). For example, a smaller sample is highly appropriate from an ethical perspective as it means that fewer subjects may have to participate in a controversial treatment.

Even with developments in software and availability of literature on the applications of optimal designs, optimal designs have not been used in observational studies as much as in experimental studies. It's important to note that both observational and experimental studies are almost equally popular in psychological research. Observational studies are studies where the assignment of subjects to groups is observed rather than manipulated. In experimental studies, the researcher has control when assigning treatments to the experimental units through randomization, thus enhancing the generalizability of the study results and at the same time avoiding potential biases. Optimal designs are particularly common in experimental designs than observational studies, probably because the researcher has control of the assignment process through randomization.

The question of potentially failing trials in longitudinal designs has been studied by several authors. Missing observations in optimal designs has been discussed by Peter Hackl (1994), Imhof, et al. (2004) and Ortega-Azurduy et al. (2008) in the context of redesigning experiments due to failing trials. In particular, Ortega-Azurduy et al. (2008) looks at the effect of missing observations on the efficiency of D-optimal designs. These studies have been focused on

the efficiency of all parameter estimates in linear mixed model (D-optimal designs). However, there is no study evaluating the loss in efficiency due to dropout if a Ds-optimal design is used in linear mixed models. The study is therefore designed to address this issue.

The rest of the paper is divided into seven sections. Section 1 is a brief introduction to Linear mixed models. Section 2 introduces characteristics of missing data. Section 3 introduces optimal designs. Section 4 covers optimal designs and response probability functions. Section 5 shows the research design of the study. Section 6 shows the results of the study and section 7 concludes the study.

SECTION 1: Linear Mixed Models

A Linear mixed model (LMM) is a statistical model with both random and fixed effects. LMMs are popular in longitudinal studies because they capture individual changes in an outcome variable using fixed and random parameters along with serially correlated errors. They are flexible and can handle dropout in the data. In addition they can model unequal number of repeated measurements from participants (Cnaan et al., 1997).

A LMM can be expressed in a matrix form as follows,

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{d}_i + \mathbf{e}_i, \quad (1)$$

$i = 1, \dots, k$, with $\mathbf{d}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Psi})$ and $\mathbf{e}_i \sim \mathbf{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i)$,

where

\mathbf{y}_i is the $q \times 1$ response vector for observations on the i th subject;

\mathbf{X}_i is the $q \times p$ design matrix for the fixed effects linking $\boldsymbol{\beta}$ to \mathbf{y}_i ;

$\boldsymbol{\beta}$ denotes a $p \times 1$ vector of fixed effects;

\mathbf{Z}_i is a known $q \times k$ design matrix for the random effects linking \mathbf{d}_i to \mathbf{y}_i ;

\mathbf{d}_i is a $k \times 1$ vector of unknown subject-specific effects;

\mathbf{e}_i is the $q \times 1$ vector of errors on the i th subject;

Ψ is the $k \times k$ covariance matrix for the random effects;

$\sigma^2 \Lambda_i$ is the $q \times q$ positive-definite error covariance matrix for the i th subject.

p represents the number of fixed parameters while the number of design points is represented by q . The vector of random effects \mathbf{d}_i represents how the i th subject deviates from the average population (Ortega-Azurduy et al., 2008). The random variables, \mathbf{e}_i and \mathbf{d}_i are assumed to be i.i.d (independent, identical and normally distributed).

SECTION 2: Missing Data

Causes and Mechanisms.

In research, missing data is as certain as taxes. Missing data can have negative effects on statistical power, and cause biased estimates if handled inappropriately (Trikriksis, 2005).

Participants could drop out of studies for various reasons: relocation, fatigue, death, or refuse to continue due to treatment failure. Careful consideration when handling missing data is important to ensure that we preserve the distributions of and the relationships among target variables. This helps reduce bias and instability in parameter estimates and standard errors. This can also help maximize statistical power.

It's worth noting that missing data patterns and rates can affect the amount of information available to estimate parameters. A low covariance coverage is an indicator for high fraction of missing information for a particular parameter. Covariance coverage is the proportion of data that is available to estimate either a correlation or covariance between two variables. Fraction of

missing information refers to the proportion of information lost due to missing data (Enders, 2010).

In addition, there is a need to understand why data are missing (missing data mechanisms). Missing data mechanisms determine which missing data technique to use. There are three missing data mechanisms (Enders, 2010) including:

- i. Missing completely at random (MCAR),

The reason for missingness is unrelated to observed variables in the model or the underlying value of the variable that is missing.

- ii. Missing at random (MAR), and

The reason for missingness is related to observed variables in the dataset but not the underlying value of the variable that is missing.

- iii. Missing not at random (MNAR).

The reason for missingness is dependent on the underlying values of the variable itself.

For the purpose of our analysis, the responses are assumed to be missing at random (MAR). The reason I choose MAR is because it can be handled by popular missing data techniques such as full information maximum likelihood and multiple imputation.

SECTION 3: Introduction to Optimal Designs

Optimal Designs

Optimal designs are a class of experimental designs that are most efficient with respect to some statistical criterion. Efficiency in this case refers to accuracy of estimators for model parameters in terms of variances of the estimators (Wong & Berger, 2009). Optimal designs

minimize the variances to achieve precise parameter estimates and improve statistical power (Kutner, et al., 2005).

Optimal designs are defined based on the optimal design theory which offers a systematic way of finding an optimal or a highly efficient design using all current information for the problem at hand (Atkinson & Donev, 1992). Smith (1918) did the original mathematical work on design experiments which is extended by Kiefer (1974) to the concept of optimal designs. Atkinson and Donev (1992) further stressed the statistical aspects of optimal designs. Optimal research designs imply that the objective of the investigation is determined before the experiment or survey is carried out and that the precision requirements for the type of data analysis are formulated, and that all possible things which could have a negative influence on the research work or could bias the results are considered (Rasch, et al, 2011).

Optimal designs can be planned prior to data collection or designs can be planned such that data collection is done sequentially or adaptively. Either way, a carefully designed study can provide accurate statistical inference with minimum cost (Wong & Berger, 2005). Wong and Berger (2009) showed that use of optimal designs can reduce number of observations from between 20% to 40% when compared to commonly used designs. This can significantly reduce budget costs without comprising accuracy of results. This study is focused only on designs planned prior to data collection.

In optimal design literature, it's a common practice to design experiments assuming that all observations are realized at the end of the study. Attrition, however, is ubiquitous in longitudinal studies. When there are missing observations, designs that assume that all observations are available at the end of the study may perform poorly (Imhof et al., 2002). Designing studies with anticipated missing observations taken into account may improve the

performance. However, researchers are more concerned about the loss in efficiency as a result of using an optimal design that assumes complete data instead of using an optimal design based on an unknown dropout process.

D(determinant)-optimality Criterion

An optimality criterion summarizes how efficient a design is and is maximized or minimized by an optimal design (Rady, et al., 2009). Optimal design studies tend to identify a design that minimizes variances of model parameter estimates, thus allowing the model to make most accurate predictions. This is achieved by use of the D criterion, which is the criterion considered in this paper due to its popularity in longitudinal design literature (Myung & Pitt, 2009). The D-criterion is also preferred as it is invariant under linear transformation of the scale of independent variables. This means that a D-optimal design remains the same with different scaling. For example, the D-optimal design in the interval $[0, 2]$ can be directly obtained by transforming the D-optimal design in the interval $[-1, 1]$. Other optimality criteria (for example A and E) are scale dependent and do not have this property (Berger & Wong, 2009).

The D-optimality criterion was first developed by Wald (1943). This criterion minimizes the area of the ellipse (two parameters) or the volume of the ellipsoid (more than two parameters) (Atkinson et al., 2007). Note that the variance-covariance matrix not only determines the shape and form of the confidence ellipsoid but also the direction of the axes.

Using a simple linear regression with one predictor variable x and an outcome variable y ($y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$) as an example, all information about the uncertainty of intercept and slope (β_0 and β_1) is contained in the confidence ellipses shown in Figure 1. The point of intersection of the dotted axes in the ellipses in Figure 1 represents the point estimates ($\hat{\beta}_0$ and $\hat{\beta}_1$) (Wong and Berger, 2009). Both the length of the axes and the volume are a measure of

uncertainty. The shorter the length of the axes and the smaller the volume of the ellipsoid, the more accurate the estimators are. In Figure 1, (i) and (ii) show a case where $\hat{\beta}_0$ and $\hat{\beta}_1$ have no correlation while (iii) and (iv) indicate a situation where the estimates are positively correlated.

D-optimality is defined as the determinant of the variance-covariance matrix of $\hat{\beta}$, $Det[Cov(\hat{\beta})]$. That is,

$$D - \text{criterion} = Det[Cov(\hat{\beta})]. \quad (2)$$

The main objective is to minimize $Det[Cov(\hat{\beta})]$, which is proportional to the inverse of the Fisher information matrix. The information matrix measures the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends. In this example, the Fisher Information Matrix is as follows:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \quad (3)$$

where \mathbf{X} represents the data matrix of independent variables. $Cov(\hat{\beta})$ is thus given by

$$cov(\hat{\beta}) = \begin{bmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} = \frac{\sigma_\epsilon^2}{N SS_x} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & N \end{bmatrix}, \quad (4)$$

where N is the sample size, SS_x is the variation of x (sum of squares of x), σ_ϵ^2 is the variance of error term. We can therefore express the D-criterion as follows:

$$D - \text{criterion} = Det[Cov(\hat{\beta})] = \frac{\sigma_\epsilon^4}{N^2 SS_x^2} [N \sum x_i^2 - (\sum x_i)^2] = \frac{\sigma_\epsilon^4}{N SS_x} \quad (5)$$

Ds-optimality

Different from D-optimal designs, Ds-optimal designs are a class of experimental designs that are efficient with respect to only a subset of parameters in a model. Ds-optimal designs are useful when the researcher is interested in a subset of parameters in a given model. When

designing for such a study it is prudent to utilize resources to estimate parameters of interest. Kiefer (1961) defined a design as Ds-optimal if it minimizes the determinant of the normalized covariance sub-matrix of estimators of the chosen model parameters while treating the other parameters as nuisance parameters.

Assuming a D-optimal model $f^T(x)\boldsymbol{\beta}$ with p parameters, it is possible to decompose f^T into f_1^T and f_2^T as follows:

$$f^T(x)\boldsymbol{\beta} = f_1^T\boldsymbol{\beta}_1 + f_2^T\boldsymbol{\beta}_2, \quad (6)$$

where $\boldsymbol{\beta}_1$ corresponds to the s parameters of interest. Thus $p - s$ parameters of $\boldsymbol{\beta}_2$ are treated as nuisance parameters. The decomposition can result in $\boldsymbol{\beta}_2$ corresponding to parameters for the blocking variables while $\boldsymbol{\beta}_1$ corresponding to the experimental factors (Atkinson et al., 2007).

Accordingly, the information matrix is also decomposed into

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}. \quad (7)$$

The covariance matrix for the least squares estimates of $\boldsymbol{\beta}_1$ is \mathbf{M}^{11} , the s by s left upper submatrix of \mathbf{M}^{-1} is

$$[\boldsymbol{\Sigma}_s(\xi)] = \mathbf{M}^{11} = [M_{11}(\xi) - M_{12}(\xi)M_{22}^{-1}(\xi)M_{12}(\xi)]^{-1} \quad (8)$$

Design ξ^* will maximize $|\boldsymbol{\Sigma}_s(\xi)|$. This is the Ds-optimum design for $\boldsymbol{\beta}_1$ and maximizes the determinant of the information matrix as follows:

$$|\boldsymbol{\Sigma}_s(\xi)| = \frac{|\mathbf{M}(\xi)|}{|M_{22}(\xi)|}. \quad (9)$$

The use of Ds-optimality designs would result in increased power since the parameters of interest are estimated more precisely (Casey et al., 2004; Atkinson & Donev, 1992).

SECTION 4: Optimal designs and response probability functions

Response probability functions.

A response probability function denotes the probability of obtaining data at a particular support point (design point). Response probability functions are used to describe the dropout process. We begin by defining a design space over which we create response probability functions.

Suppose ξ is a design in the design space Δ containing all designs of interest: i.e.

$$\xi = \left\{ \begin{matrix} x_1 & x_2 & x_3 & \dots & x_d & \dots & x_{s-1} & x_s \\ w_1 & w_2 & w_3 & \dots & w_d & \dots & w_{s-1} & w_s \end{matrix} \right\}. \quad (10)$$

There are m design points, x_1 - x_s distinct values of the independent variable (IV), x . w_1 - w_s are the associated weights, where $w_i = \frac{n_i}{N}$. Instead of using weights, we can also use sample sizes n_1 - n_s for discrete designs. An illustration using dosage example is provided. Suppose there are 6 different dosage levels on the design interval $\Delta = [1, 6]$ and a sample size of $N=60$ patients. If we have equal proportion of patients assigned at each dosage level, all the weights are equal to $w_s = \frac{1}{6}$ and each dosage level is $n_i \times \frac{1}{60} = \frac{10}{60}$. To facilitate interpretation and comparison of different designs, the design points are recoded between -1 and 1 by the transformation $d'_j =$

$\frac{d_j - \tilde{d}}{6 - \tilde{d}}$, where $\tilde{d} = 1 + (6 - 1)/2$. The resulted design is

$$\xi_{60} = \left\{ \begin{matrix} -1 & -.6 & -.2 & .2 & .6 & 1 \\ \frac{10}{60} & \frac{10}{60} & \frac{10}{60} & \frac{10}{60} & \frac{10}{60} & \frac{10}{60} \end{matrix} \right\}. \quad (11)$$

Let's assume that $p(x_d)$ is the response probability of measuring a subject at time points x_d . For the design point x_1 the response probability is $p(x_1)$ with a sample size n_1 , while for the d^{th} design point (x_d), the response probability is $p(x_d)$ with a sample size n_d : ($n_d = n_1 p(x_d)$).

Several assumptions are made during this process of determining response probability functions.

These include:

- 1) $n_s \geq 1$ i.e., at least one subject is observed at the last point.
- 2) $p(x_d)$ is a monotonically decreasing function with
 $p(x_1) \geq p(x_2) \geq \dots \geq p(x_{s-1}) \dots \geq p(x_s)$, where $p(x_1) > p(x_s)$.
- 3) Dropout occurs in a non-informative mechanism (MAR).

Optimal designs with complete data

The D-optimality (which is our optimality criterion) minimizes the generalized variance (or the determinant of the variance covariance matrix) of the parameter estimates $\hat{\beta}$ (Atkinson & Donev, 1996). We define a design ξ^* as D-Optimal for model Q if:

$$Det[Var_Q(\hat{\beta}|\xi^*)] \leq Det[Var_Q(\hat{\beta}|\xi)] \quad \forall \xi \in \Delta, \quad (12)$$

where Q is a specific model from a set of LMMs. Minimizing this determinant maximizes the power of the simultaneous test of the null hypotheses for the parameters in vector β .

Optimal designs with incomplete data

While with complete data, we have a total of s responses (complete responses from all design points), with missing data, we will have $d < s$ responses. The response probability $p(x_d)$, and the matrices for fixed effects $\mathbf{X}^{[d]}$ and random effects $\mathbf{Z}^{[d]}$ in a linear mixed model is $P(x_d) = n_d / n_I$.

The matrices $\mathbf{X}^{[d]}$, $\mathbf{Z}^{[d]}$ are of size $d \times p$ and $d \times k$ respectively, and their corresponding covariance matrix is:

$$\mathbf{V}^{[d]} = \mathbf{Z}^{[d]} \boldsymbol{\Psi} \mathbf{Z}'^{[d]} + \sigma^2 \mathbf{\Lambda}_I \quad (13)$$

The covariance matrix of the best linear unbiased estimator $\hat{\beta}$ can be then written as

$$Var(\hat{\beta}) = (\sum_{d=1}^s m_d \mathbf{X}^{[d]'} \mathbf{V}^{[d]-1} \mathbf{X}^{[d]})^{-1} \quad (14)$$

One of the challenges of finding an optimal design for a linear mixed model with dropout is that the response probability function is unknown at the design stage of study. However, prior knowledge on potential failing trials in the area (from past research or pilot studies) can be incorporated in the design.

SECTION 5: Method

Motivation Example

An example model would be a LMM investigating the effect of different interventions for treating children with attention-deficit/hyperactivity disorder (ADHD). This would be a popular model in clinical psychology. Some of the available interventions include medication (treatment A) and intensive behavioral treatment (treatment B). Studies have shown that in the long run, behavioral treatments are highly effective (Fabiano et.al. 2008) and that there was no long term side effects. Suppose that the outcome variable is academic performance. Since medication does not directly address problems related to children's academic performance, only the effect of behavioral treatment on academic performance is considered and we will utilize a subset of variables in our model. This is a Ds-optimal model which provides a more precise estimate of the effect of the treatment B. The study is conducted in the context of the example.

The LMM is expressed in equation (1) above. Using the example given above, we show the equations involved in determining the subset of variables of interest for the Ds-optimal model. There are three treatment groups (intensive behavioral-A, medication-B, and control-C) which were measured in zero, nine and fourteen months. The LMM can be written as follows:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \epsilon_{ij} \quad (15)$$

where

Y_{ij} is the response of i^{th} subject, j^{th} occasion.

β_0 is the average response at start of treatment

$\beta_1, \beta_2, \beta_3$ represent average time effects for treatment A, B and Control groups

L_i, H_i, C_i are indicator variables defined to be 1 if the child belongs to A, B and C groups respectively

b_{1i} represents subject specific random effects

t_{ij} represents time since baseline

$(\beta_0 + b_{1i})$ represents subject specific mean response at baseline.

Models for each treatment group are as follows:

$$Y_{ij} = \begin{cases} \beta_0 + b_{1i} + (\beta_1 + b_{2i})t_{ij} + \epsilon_{ij}, & \text{treatment A} \\ \beta_0 + b_{1i} + (\beta_2 + b_{2i})t_{ij} + \epsilon_{ij}, & \text{treatment B} \\ \beta_0 + b_{1i} + (\beta_3 + b_{2i})t_{ij} + \epsilon_{ij}, & \text{Control} \end{cases} \quad (16)$$

Since the medication intervention does not necessarily address problems related to a child's academic performance, β_2 is excluded from the model which results in a Ds-optimal model as shown below:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_3 C_i + b_{2i})t_{ij} + \epsilon_{ij} \quad (17)$$

Design Evaluation

Recall that the goal of this study is to examine the effect of drop out on the efficiency of D-optimal (denoted by ξ_d^*) and Ds-optimal designs (denoted by ξ_{ds}^*) for LMMs. A design ξ^* is defined as Ds-optimal design for model Q if:

$$Det[Var_Q(\hat{\beta}|\xi^*)] \leq Det[Var_Q(\hat{\beta}|\xi)] \quad \forall \xi \in \Delta \quad (18)$$

The probability of observing all subjects at all design points is less than one, i.e., $p(x_d^*) <$

1 for all d . The response probability is given by $(x_d^*) = \frac{n_d}{n_1}$, $n_1 > n_2 > \dots > n_d \dots > n_{s-1} >$

n_s . The aim here is to evaluate loss of efficiency, by comparing the efficiency of ξ_d^* to ξ_{ds}^* . Ds efficiency is the ratio of the variance of the subset of parameters of interest under the Ds-optimal design, ξ_{ds}^* , to the variance of the same subset of parameters under the D optimal design, ξ_d^* .

$$RE(\xi_d^* | \xi_{ds}^*) = \left[\frac{Det[VAR_Q(\hat{\beta} | \xi_{ds}^*)]}{Det[VAR_Q(\hat{\beta} | \xi_d^*)]} \right]^{\frac{1}{p}} \quad (19)$$

The p^{th} root enables us to determine the number of replicates of an optimal design ξ_d^* that will be as efficient as one replicate of the optimal design ξ_{ds}^* . In terms of sample size, RE enables us to determine the number of extra participants needed for ξ_d^* to have the same efficiency as ξ_{ds}^* . For instance, if $RE(\xi_d^* | \xi_{ds}^*) = 0.65$, then we need $(.65^{-1} - 1) * 100\% = 54\%$ more subjects for ξ_d^* to have same efficiency like ξ_{ds}^* .

Generate Missing Data

We utilize both a quadratic and a linear response function to capture the dropout process. The amount of missingness increases from 0% at the first design point to 70% at the last design point.

$$P_{lin}(X_j) = 0.65 - 0.35X_j \text{ . and } P_{quad}(X_j) = 0.5 - 0.35X_j + 0.15X_j^2 \quad (20)$$

where X_j are the design(support) points of longitudinal observations Y for a sample of subjects.

The software used in this study is R version 3.1.3. The study hypothesis is that when drop out is taken into account, D-optimal designs are less efficient relative to Ds-optimal designs. Ds-optimal designs are more efficient because the models are simpler compared to D-optimal design which gives more degrees of freedom (information) to estimate the variance more precisely.

Design Factors

Four classes of LMMs with different variance-covariance matrix \mathbf{T} will be investigated as shown below:

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix} \quad (21)$$

- i) Fixed effects models with no random effects
- ii) random intercept model
- iii) random slope and intercept model where there is no covariance between intercept and slope
- iv) a random intercept and slope with covariances between random intercept and slope equal to $t_{12} = \rho^* \sqrt{t_{11} x t_{22}}$. Correlation between the random intercept and slope is 0.8.

For these four classes, the elements of \mathbf{T} are varied at two levels as follows: $t_{11}=\{0,1\}$, $t_{22}=\{0,3\}$, and $t_{12} = \{0,0.8\sqrt{3}\}$. This choice is made because past research (Ortega-Azurduy et al., 2008) has shown that trends of the optimal allocation of the time points with drop out depended more on the form of the variance-covariance matrix \mathbf{T} than on specific values of its individual elements t_{11} , t_{22} and t_{12} . In this study I consider $p = 2, 3$ and 4 fixed polynomial parameters with $q=3$ and 4 time points. From the four classes of LMMs investigated, we note that class 4 portrays the core characteristics of a longitudinal study. Class 4 has both fixed and random effects and a correlation between the random intercept and slope.

For both D and Ds-optimal designs, I will examine the effect of missing observations using the response function and then evaluate the efficiency loss using relative efficiency plots. Plots will be used to show relative efficiency for the various models within the design space given the response functions.

SECTION 6: Results

D and Ds-optimal designs for data with dropout

D-optimal and Ds-optimal designs (D and Ds) are compared for number of fixed parameters 2 and 3, and number of design points 3 and 4. Figures 3 – 6 display D and Ds with complete and data with missing observations. Each figure has 3 columns and 4 rows. The three columns compare the D and Ds designs for data that is complete, data having a linear dropout and data having a quadratic dropout function, respectively. For each of the 12 plots in a figure, the x-axis represents the support points, x_j , with range between -1 and 1. The correlation parameter is displayed on the-axis and ranges from 0.1 to 0.8. The four rows in each figure are relate to the four classes of the variance covariance matrix \mathbf{T} of the random parameters.

Figures 3 and 4 display a first degree polynomial (i.e., linear model). For both D and Ds-optimal designs with dropout, in all classes, there is more displacement with the quadratic dropout function than the linear function when compared to complete data. Displacement varies with the level of the correlation parameter. The direction of displacement is not dependent of dropout function. Most of the design points with dropout shift to the left in comparison to the design points for complete data.

Figures 5 and 6 display a second degree polynomial (i.e., quadratic model). The displacement significantly reduces for the quadratic model for all designs for data with dropout irrespective of the variance and covariance matrix of the random parameters. The difference in displacement between the D and Ds designs is decreased too. An example is figure 3 (class 3) and figure 5 (class 3). The rate of displacement of the design points (time points) decreased as the correlation parameter reduces from 0.8 towards 0.1. An example is Figure 6 classes 4.

Efficiency loss – comparison of the designs with dropout to their complete data counterparts

Figure 7 compares the relative efficiency of Ds-optimal designs to D-optimal designs for the linear and quadratic response functions. There are six plots with each plot showing a different combination of time point and number of fixed parameters. The x-axis represents the models M_ω where $\omega = 1$ to 20. A list of the models' elements of the variance and covariance matrix \mathbf{T} is shown in table 1. Relative Efficiency (RE) is on the y axis. There is a threshold line drawn at RE = 0.85.

For a linear mixed effect model with three and four design points (Figure 7, row 1), the efficiency is at least 0.87. The linear and quadratic dropout functions indicate that the highest loss of efficiency is observed as p approaches zero (< 0.2). This can be observed in models 1,2,16, and 17. Figure 7 row 2 shows the quadratic model with three and four design points in column 1 and 2, respectively.

The quadratic model with three design points seems to have more loss in efficiency compared to the quadratic model with four time points. For the quadratic model with three time points, model 6 has a RE of 0.85, while 1, 2, 3, and 7 have RE of 0.86. The quadratic models with four time points have most models having a RE of at least 0.98. Models 15, 19 and 20 seem to have RE of around 0.90.

To investigate the change in efficiency as parameters and time points increase, I included a cubic models with four and five time points. For the cubic models with four time points, the lowest RE was around 0.87. For the cubic models with five time points, the lowest RE was around 0.97. The two columns of plots in Figure 7 look basically the same, showing that relative efficiency (RE) is not depended on the particular dropout function.

As the number of parameters increase past three, both designs perform almost the same. The autocorrelation parameter seems to play a significant role in determining RE. The models showing significant loss in efficiency have low values (< 0.3) of the autocorrelation parameter. Besides the autocorrelation parameter, other factors that are not statistical can affect the design of choice. These include (and not limited to) ethical and logistical complexities, trial safety, and budgetary constraints. Even though a linear and a quadratic response function is used in this analysis, past research shows that different drop out functions result to similar small efficiency losses. For all our models, the RE was at least 0.85. This implies that with a dropout of up to 70%, the loss in efficiency did not exceed 15%.

SECTION 7: Conclusions

This paper evaluates the change in efficiency when D and Ds-optimal designs with dropout are compared with D-optimal designs with complete data. Past research has shown that with drop out, an optimal design that assumes complete data does not remain optimal. We note that from an efficiency stand point, both D and Ds-optimal designs are robust to dropout.

Shifting of the time points (as a result of dropout) for the optimal designs is biggest for LMMs having two parameters (linear). This may be due to the amount of information lost due to missing data. For a linear model with two parameters, if the slope parameter was the parameter of interest, the Ds-optimal design would have one parameter, so drop out would imply significant loss of information leading to huge displacement of the time points. For LMMs with more than two parameters, shifting/displacement of time points is reduced. For D and Ds-optimal designs with dropout, the shifting of the design points increases as the correlation parameter increases. This may be attributed to the fact that at high levels (> 0.6) of the correlation parameter, any dropout will result to a greater loss of information than at a lower level (towards zero). For the

two parameter model with 4 time points, the Ds-optimal design with complete data is closer to the Ds-optimal design with dropout (for both dropout functions) compared to the D-optimal with complete data and the D-optimal with dropout. This means that the determinant of the variance and covariance matrix for the Ds-optimal design with dropout is closer in size to that of Ds-optimal design with complete data compared to the D-optimal design (with complete data and with dropout).

From the efficiency plots, we observe that as the number of time points increase, both the D and Ds-optimal designs lose less efficiency. This is because additional measurements (time points) would provide more information which improves precision of the parameter estimates, and the model.

Whenever there is a shift in the time points for any design, the researcher needs to respond appropriately to counter the effect of the dropout in successive time points. If for example the optimal design points are shifting to the left compared to those for the design with complete data, it will be prudent for the researcher to check where the variation of the displacement of the design points is largest within the study period. More measurements need to be taken in the period of the study with the largest variation.

Also, in cases where we have a high autocorrelation parameter (above 0.6), follow-up evaluations should be taken at the section with the highest response probability. This is beneficial as the higher the response probability, the more information we have.

Although dropout is a reality in longitudinal studies, the use of a D-optimal design results in loss of efficiency not exceeding 15% when we experience dropout of up to 70%. Besides reducing the number of time points and participants (which brings down cost and may have

ethical benefits), longitudinal studies can benefit from the use of optimal designs as they are robust to dropout.

When a researcher is interested in a subset of parameters in a model, a Ds-optimal design gives more precise results when compared to a D-optimal design. However, if the autocorrelation exceeds 0.4, the difference in precision in terms of efficiency seems to be negligible. Researchers have a choice to increase their sample size to compensate for the loss in efficiency.

This study assumes a first order auto-regressive covariance structure. Future research may benefit from looking at other correlation structures. From a Hierarchical Linear Model (HLM) perspective, it may be useful to study how higher level predictors (for example, level-two units) affect efficiency when there is dropout.

References

- Atkinson, A. C., Donev, A. N. (1992). Optimum Experimental Designs. Oxford Science Publications.
- Atkinson, A. C., Donev, A. N., (1996) *Optimum Experimental Designs*. Oxford Science Publications: Oxford.
- Atkinson, A. C., Donev, A. N. and Tobias R. D., (2007). Optimum Experimental Designs, with SAS Oxford, Oxford University Press
- Berger, M. P. F. (1994). D-Optimal Sequential Sampling Designs for Item Response Theory Models. *Journal of Educational Statistics*, 19, 43-56.
- Berger, M. P. F., Wong, W. K. (2005). Applied Optimal Designs. John Wiley and Sons.
- Berger, M. P. F., Wong, W. K. (2009). An Introduction to Optimal Designs for Social and Biomedical Research. John Wiley and Sons.
- Casey M., Gennings C., Carter W.H., Moser C. W., & Simmons E.J., (2004). Ds-optimal studies for studying combinations of chemicals using multiple fixed –ratio experiments.
- Cnaan A, Laird N.M., Slasor P. (1997). Tutorial in biostatistics: using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*; **16**:2349–2380.
- Cox, D.R. (1958). Planning of Experiments. New York: John Wiley & Sons, pp. 141-142
- Enders K. C. (2010). Applied Missing Data Analysis. The Guilford Press.
- Fabiano G. A., Pelham W.E., Coles E.K., Gnagy E.M., Chronis-Tuscano A., O'Connor B.C. (2008). A meta-analysis of behavioral treatments for attention-deficit/hyperactivity disorder. *Clinical Psychology review*; 29, 129-140.
- Hackl, Peter (1994). Optimal Design for Experiments with Potentially Failing Trials.
- Haines L.M., Perevozskaya I., Rosenberger W.F. (2003). Bayesian optimal designs for Phase I clinical trials. *Biometrics*; 59(3): 591–600.
- Holling, Heinz (2013). Current Issues in Optimal Design *Zeitschrift fur Psychologie* 2013; Vol. 221(3):121–123.
- Imhof LA, Song D, Wong WK. (2002). Optimal design of experiments with possibly failing trials. *Statistica Sinica*; 12:1145–1155.
- Imhof, L., Song, D., Wong, W. K. (2004). Optimal Design of Experiments with Anticipated Pattern of Missing Observations. *Journal of Theoretical Biology*, 228, 251-260.

- Myung J.J., Pitt M.A. (2009). Optimal Experimental Design for Model Discrimination
Psychological Review, Vol 116(3), Jul 2009, 499-518.
- Kiefer, J. (1961). Optimum Designs in Regression Problems II. *Annals of Mathematical Statistics*, 32, 298-325.
- Kiefer, J. (1974). General Equivalence Theory for Optimum Designs (Approximate Theory).
The Annals of Statistics, 2, 849-879.
- Kutner M.H.,Nachtsheim C.J.,Neter J.,Li W. (2005). *Applied Statistical Models*, 5th edn.
McGraw-Hill/Irwin series Operations & decision sciences.
- Maus, B., & van Breukelen, G.J.P. (2013). Optimal design for functional magnetic resonance
imaging experiments: Methodology, challenges and future perspectives. *Zeitschrift für
Psychologie*, 221, 174–189.
- McClelland G.H.(1997).Optimal Design in Psychological Research. *Psychological Methods*, Vol
2, No 1, 3-19.
- Moerbeek, M., & Joz'wiak, K. (2013). Optimal designs for event history analysis. *Zeitschrift für
Psychologie*, 221, 160–173.
- Mullis, I. V., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y.,& Preuschoff, C. (2009). TIMSS
2011 assessment frameworks. Chestnut Hill, MA: TIMSS & PIRLS International Study
Center
- Ortega-Arzuduy, S. A., Tan, F.E.S., & Berger, M.P.F.(2008). The effect of drop out on the
efficiency of D-optimal designs of linear mixed models. *Statistics in medicine*, 27,2601-
2617.
- Rady A.E.,et al (2009) Relationships among Several Optimality Criteria
<http://interstat.statjournals.net>
- Rasch, D.,Pilz,J.,Verdooren,R.,Gebhart,A.(2011). *Optimal Experimental Design with R*. CRC
Press
- Smith, K. (1918), On the Standard Deviations of Adjusted and Interpolated Values of an
Observed Polynomial Function and its Constants and the Guidance They Give
Towards a Proper Choice of the Distribution of Observations, *Biometrika*, 12, 1-85.
- Tsikriktsis N.(2005). A review of techniques for treating missing data in OM survey research. *Journal of
Operations Management* 24, pp 53-62.
- Wald, A. (1943). On the efficient design of statistical investigations. *Annals of Mathematical
Statistics*, 14, 134–140.

Figures

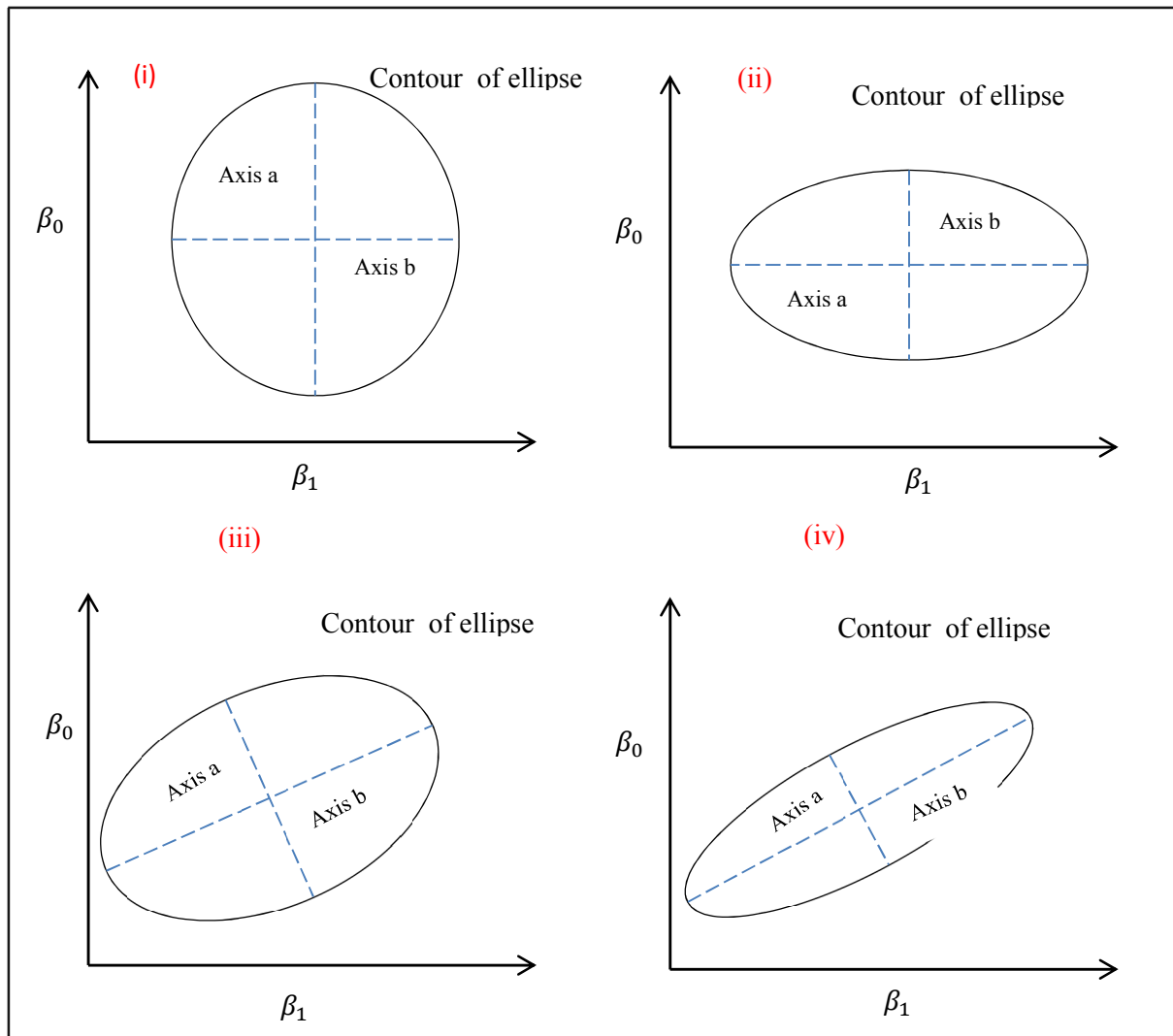


Figure 1. Confidence ellipses for two parameters β_0 and β_1 in the simple linear model.

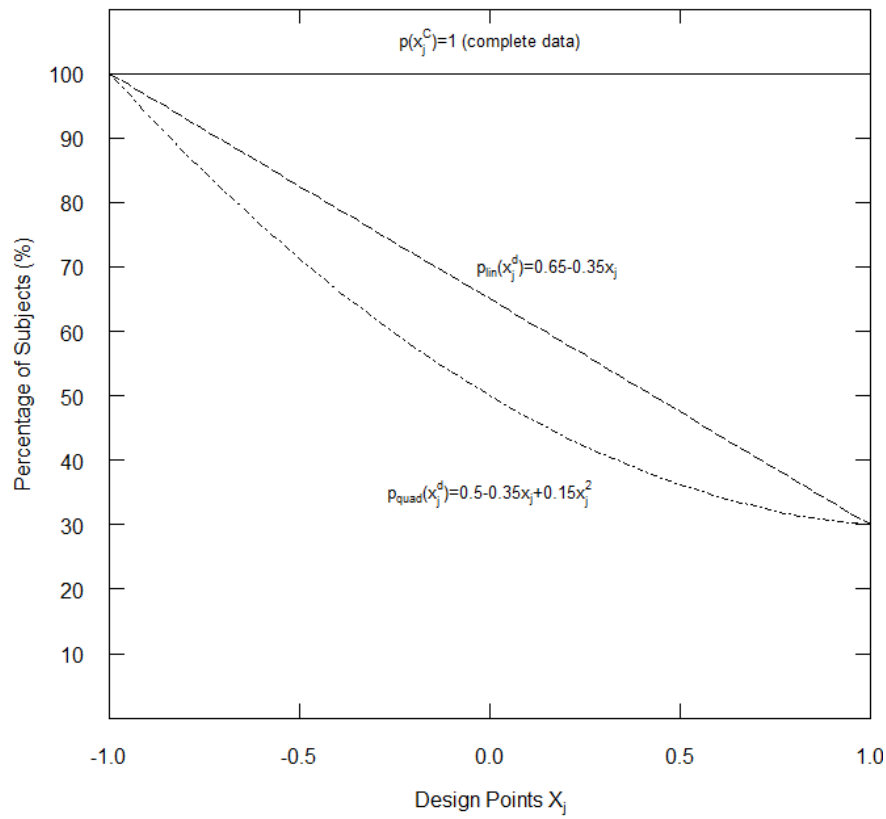


Figure 2. Drop out probability functions (p_x): $p_{lin}(x_j^d)$ – *linear*, $p_{quad}(x_j^d)$ – *quadratic*, $p(x_j^c)$ – *complete data*

D_optimal vs Ds_optimal Design for $p = 2$, $q = 3$

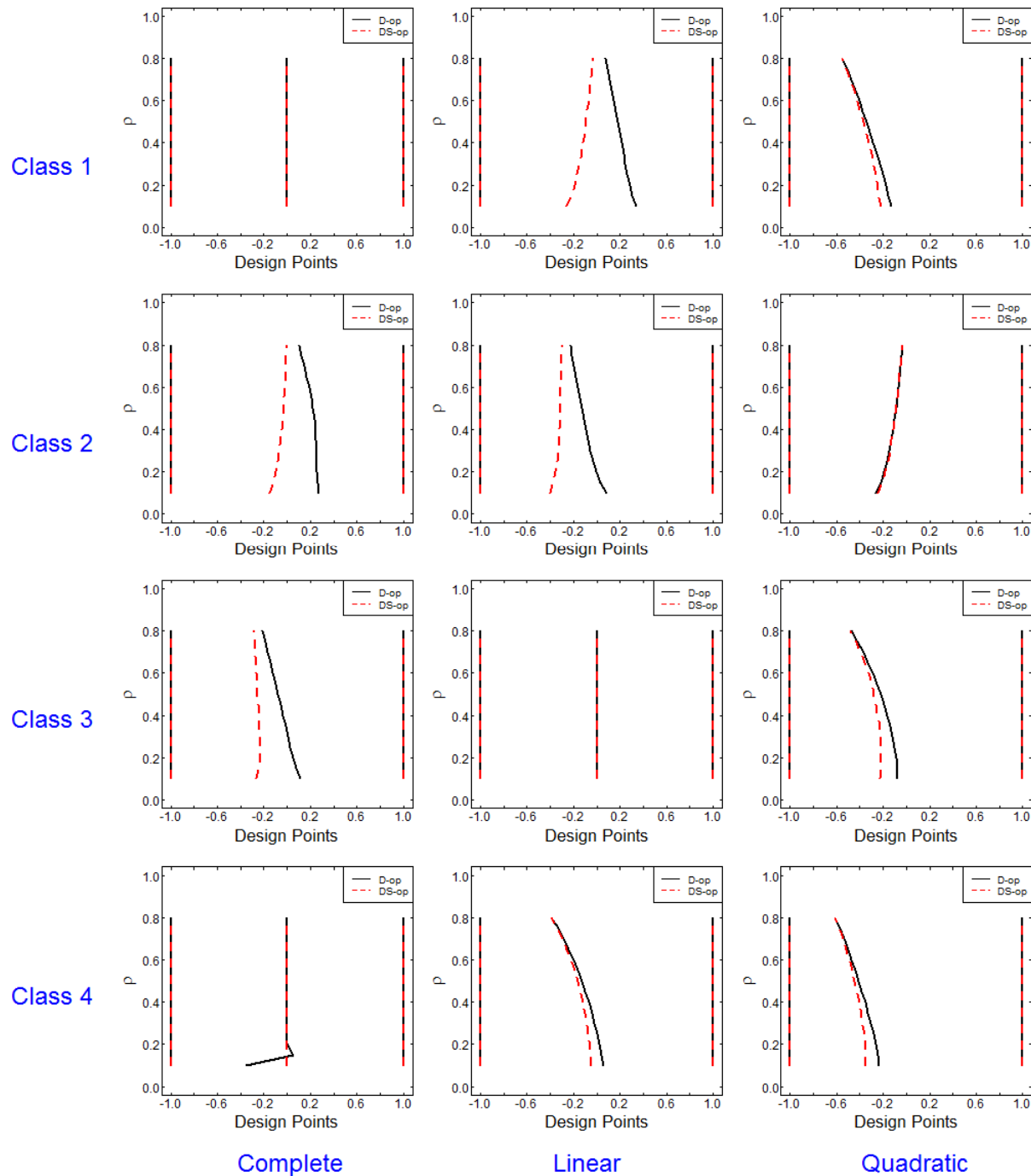


Figure 3 - D and Ds-optimal design for a first degree polynomial, with 3 time points

D_optimal vs Ds_optimal Design for $p = 2$, $q = 4$

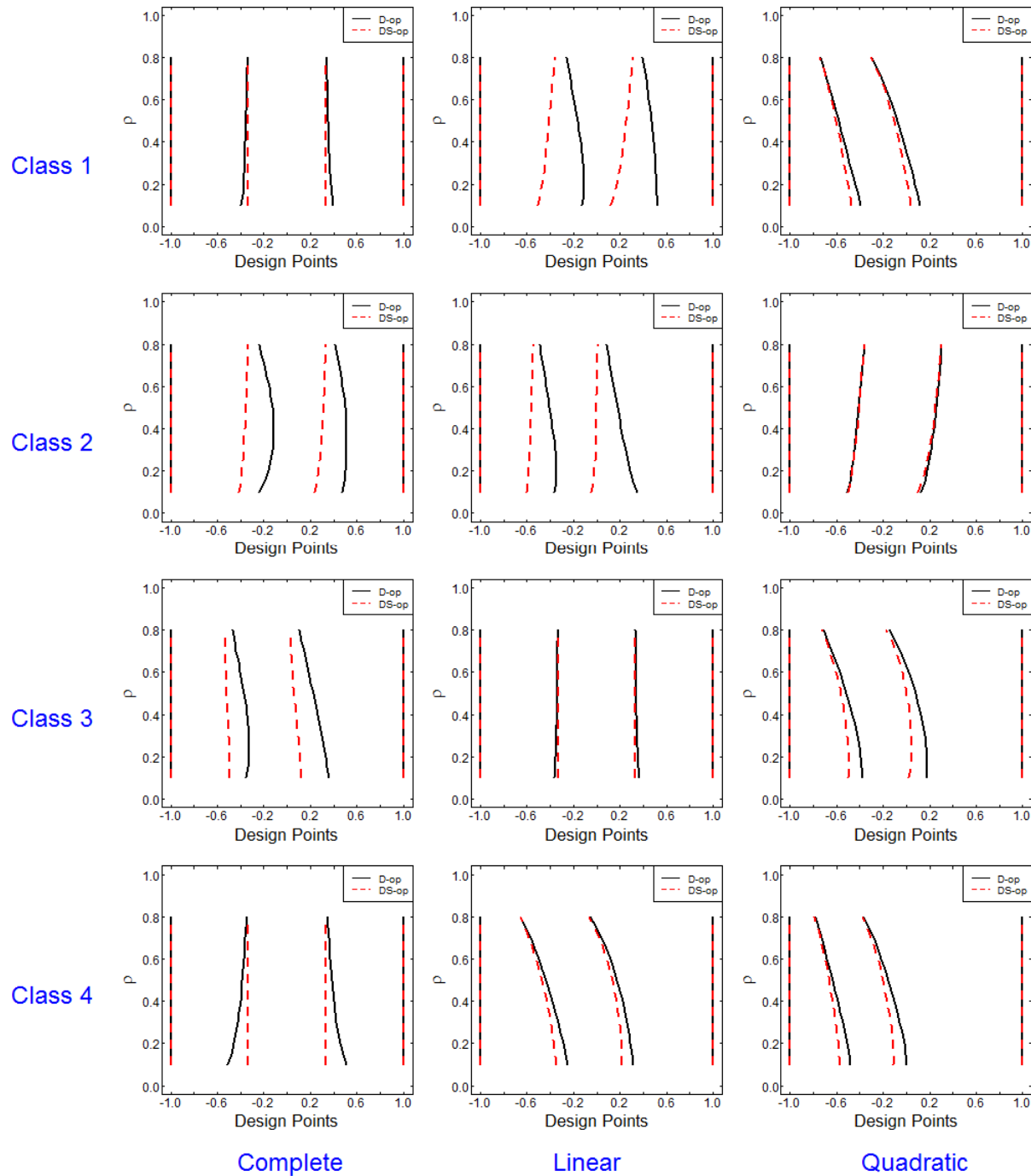


Figure 4 - D and Ds-optimal design for a first degree polynomial, with 4 time points

D_optimal vs Ds_optimal Design for $p = 3$, $q = 3$

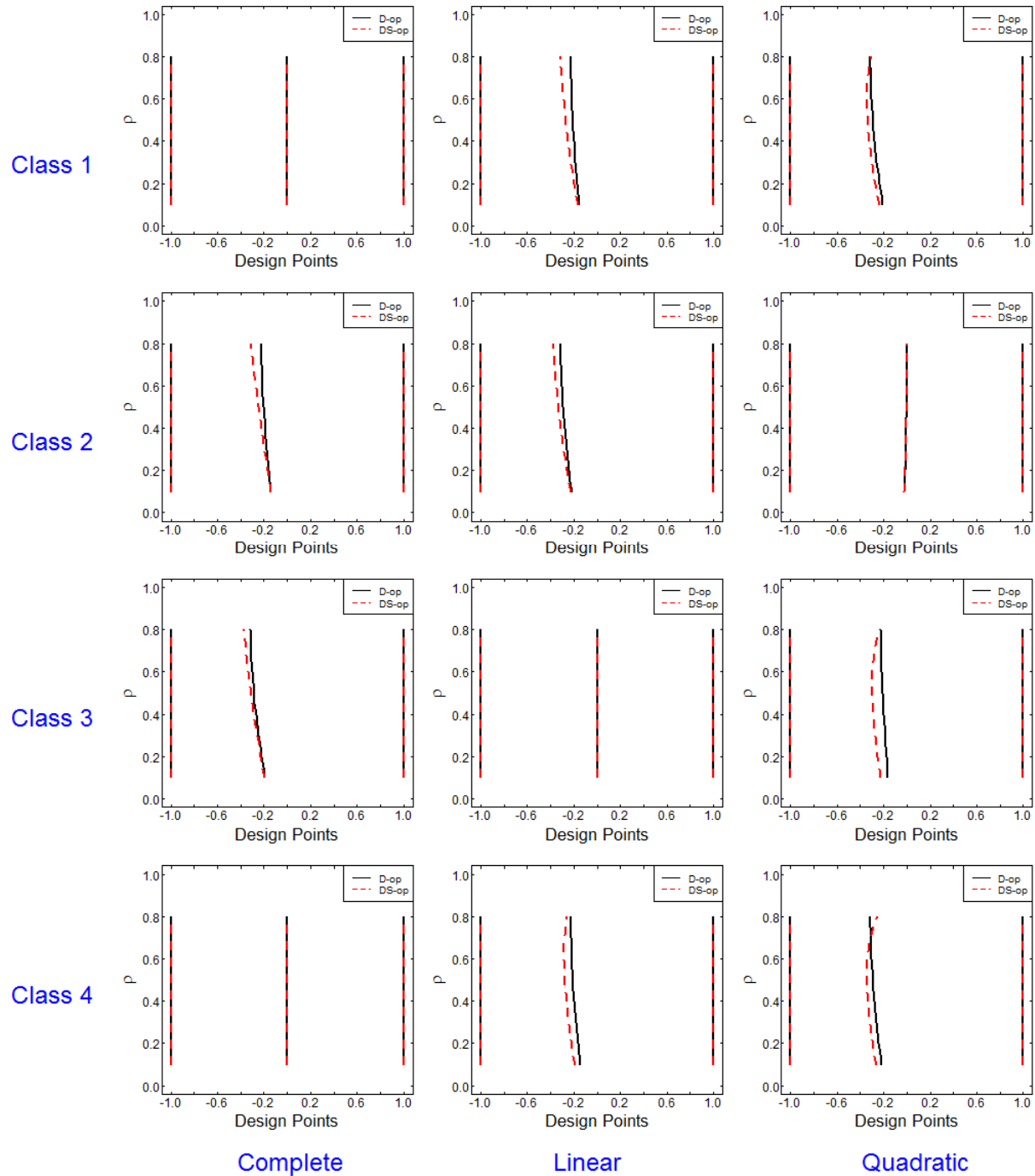


Figure 5 - D and Ds-optimal design for a second degree polynomial, with 3 time points

D_optimal vs Ds_optimal Design for $p = 3$, $q = 4$

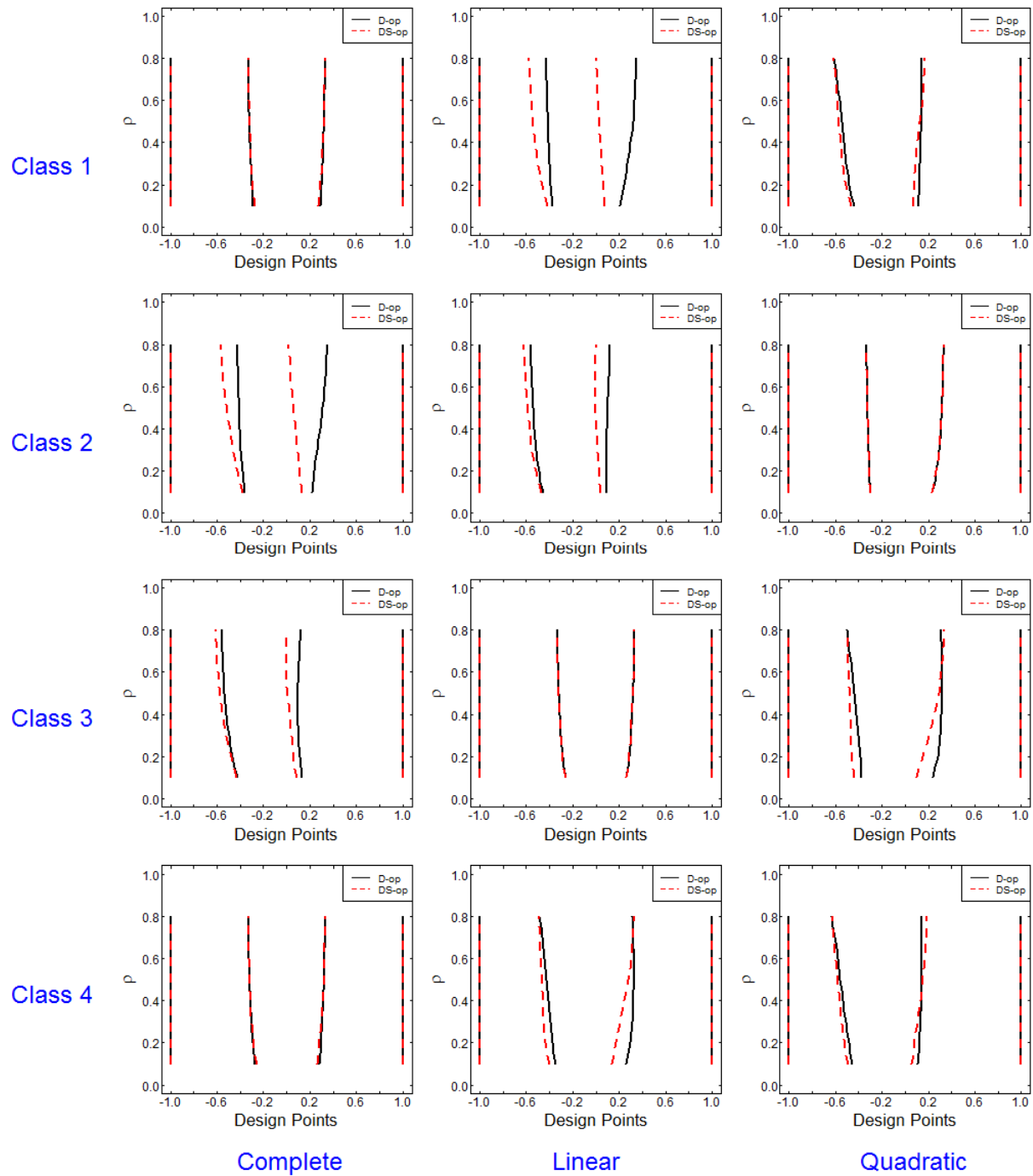


Figure 6 - D and Ds-optimal design for a second degree polynomial, with 4 time points

Relative Efficiency (RE)

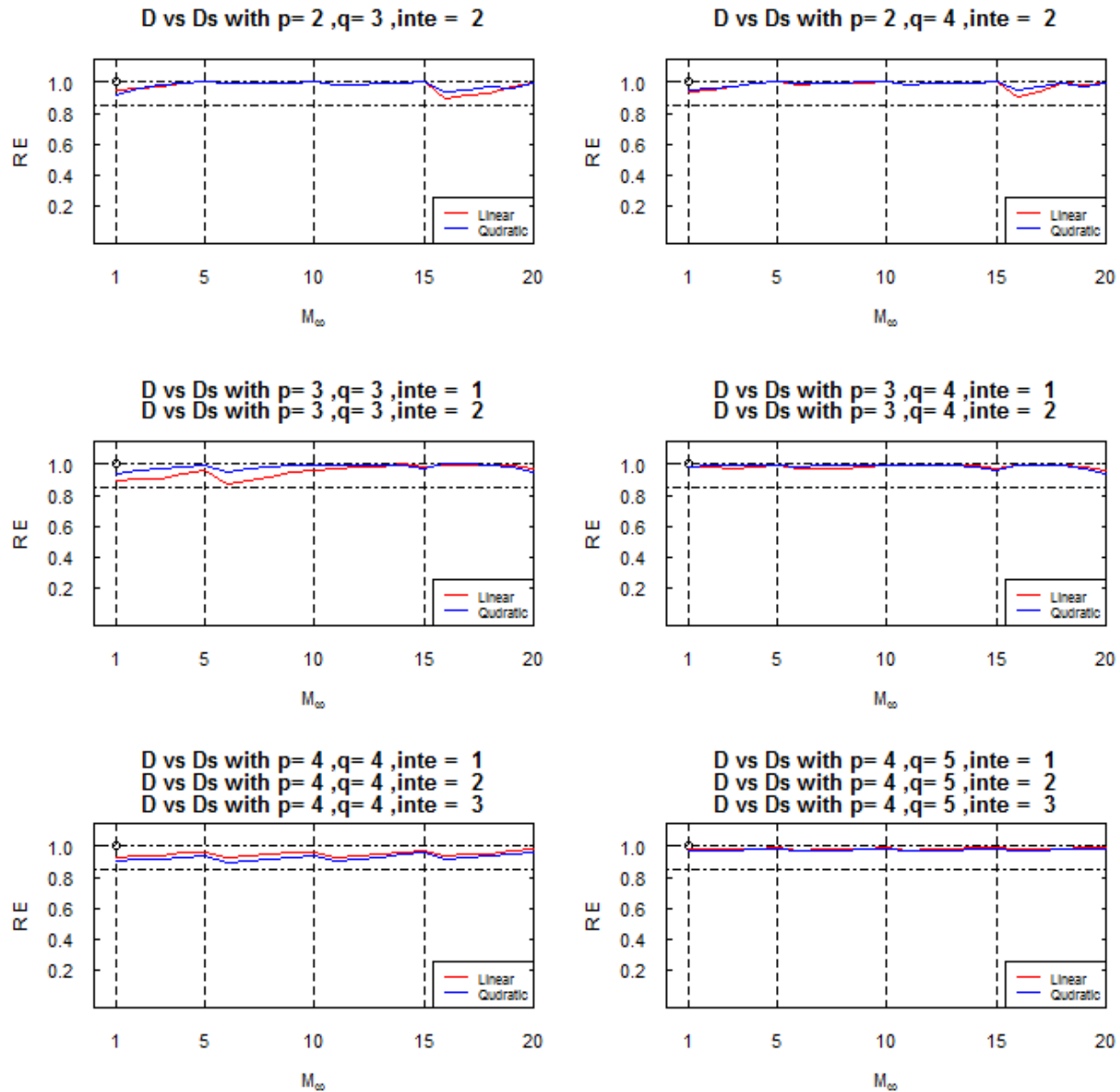


Figure 7 - Relative efficiency - D and Ds-optimal designs

Row1 - comparing D and Ds-optimal designs for the linear and quadratic response functions. For the Ds-optimal designs, the parameter of interest is β_2 .

Row2 - comparing D and Ds-optimal designs for the linear and quadratic response functions. For the Ds-optimal designs, the parameter of interest is β_1 and β_3 .

Row3 - comparing D and Ds-optimal designs for the linear and quadratic response functions. For the Ds-optimal designs, the parameter of interest is β_1 , β_2 , and β_3 .

Table 1

Models Evaluated under the Four Classes of LMMs

| Fixed Effects | | | | | | Random slope & intercept | | | | | |
|-----------------------|-----|-----|-----|-----|-----|---|-----|-----|------|------|-----|
| model | rho | t11 | t12 | t21 | t22 | model | rho | t11 | t12 | t21 | t22 |
| 1 | 0.1 | 0 | 0 | 0 | 0 | 11 | 0.1 | 1 | 0 | 0 | 3 |
| 2 | 0.2 | 0 | 0 | 0 | 0 | 12 | 0.2 | 1 | 0 | 0 | 3 |
| 3 | 0.3 | 0 | 0 | 0 | 0 | 13 | 0.3 | 1 | 0 | 0 | 3 |
| 4 | 0.6 | 0 | 0 | 0 | 0 | 14 | 0.6 | 1 | 0 | 0 | 3 |
| 5 | 0.8 | 0 | 0 | 0 | 0 | 15 | 0.8 | 1 | 0 | 0 | 3 |
| Random intercept only | | | | | | Random intercept, slope with covariance | | | | | |
| model | rho | t11 | t12 | t21 | t22 | model | rho | t11 | t12 | t21 | t22 |
| 6 | 0.1 | 1 | 0 | 0 | 0 | 16 | 0.1 | 1 | 1.39 | 1.39 | 3 |
| 7 | 0.2 | 1 | 0 | 0 | 0 | 17 | 0.2 | 1 | 1.39 | 1.39 | 3 |
| 8 | 0.3 | 1 | 0 | 0 | 0 | 18 | 0.3 | 1 | 1.39 | 1.39 | 3 |
| 9 | 0.6 | 1 | 0 | 0 | 0 | 19 | 0.6 | 1 | 1.39 | 1.39 | 3 |
| 10 | 0.8 | 1 | 0 | 0 | 0 | 20 | 0.8 | 1 | 1.39 | 1.39 | 3 |